DOCUMENT RESUME

ED 090 272 TM 003 548

AUTHOR Rippey, Robert M.
TITLE Confidence Test Scoring and Incentive Conditions.
PUB DATE [Apr 74]
NOTE 14p.; Paper presented at the National Council on Measurement in Education Annual Meeting (Chicago, Illinois, April 16-18, 1974)

EDRS PRICE MF-$0.75 HC-$1.50 PLUS POSTAGE
DESCRIPTORS *Confidence Testing; *Guessing (Tests); High School Students; *Motivation; Scoring; Sex Differences; Social Differences; *Testing; Test Reliability; Test Results; Test Wiseness

ABSTRACT

The effects of incentive conditions on the results of a confidence test were investigated. Two hundred thirty high school subjects were administered a very difficult confidence scored test under two conditions: 1) that the test would count heavily on their grades (incentive condition) and 2) that the test was for research purposes and would not be counted (relaxed condition). An analysis of the data revealed: 1) Under incentive conditions, scores on confidence tests are higher, and reliability significantly lower when compared to the relaxed conditions. 2) Females have a greater tendency toward taking extreme positions than males, especially in the incentive condition. 3) Subjects in the incentive group liked the test better, had more of a tendency to take extreme positions, and made more appropriate estimates of their degree of confidence. 4) Middle socioeconomic subjects, compared to both upper and lower socioeconomic subjects, made higher scores and more appropriate estimates of confidence. 5) High scoring subjects gambled more on difficult items under the relaxed condition, but gambled less on difficult items in the incentive condition. 6) Positive attitudes toward the tests were directly related to degree of confidence. (Author/MLP)

10.14

CONFIDENCE TEST SCORING AND INCENTIVE CONDITIONS

A Paper Presented to the NCME
April 16, 1974
Chicago, Illinois

Robert M. Rippey, Professor
University of Connecticut
Schools of Medicine & Dental Medicine
Farmington, Connecticut 06032

U.S. DEPARTMENT OF HEALTH, EDUCATION & WELFARE
NATIONAL INSTITUTE OF EDUCATION
THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL NATIONAL INSTITUTE OF EDUCATION POSITION OR POLICY

ED 090272

Confidence testing asks subjects to assign probabilities of confidence to the options of multiple choice items. Considerable disputation has arisen over the importance and efficacy of these procedures (Hambelton, Roberts, and Traub, 1970) and Rippey (1970), and a summary of some of the arguments is contained in Wang and Stanley (1970). My own continued interest in confidence testing lies not in the area of the alleged improved psychometric properties of confidence tests, but in the area outlined rather early by DeFinetti. How do we get persons to become better assessors of their own confidence?

Accurate assessments of confidence are especially important in areas involving incomplete knowledge of data, and in areas where important decisions must be based on an inadequate body of theory. Some of the early work in confidence testing was based on utility theory. Scoring functions were developed which produced maximum scores in the long run if and only if the subject maximized his expected utility, given a knowledge of the payoff of his choices (Shuford, Albert, and Massengill, 1966). Unfortunately, one man's utility is sometimes another man's poison. There are differences in sex, social class, and condition of administration which interact with item difficulty and contribute to error variance in the confidence testing situation.

The Meaning of Confidence

Two hundred sixty-three sophomore and junior students from a high school in a suburb of Chicago were randomly assigned to two groups and administered fifteen very difficult items from the STEP Writing Test, Level 1. Ss were told that the items might or might not have unique correct responses.[1] One group was told that the test they were taking would count toward their grades in English. The other group was told that the test was being administered for research purposes and would not be counted on their grades. The teachers were given the grades of the subjects in the incentive group, and they had agreed to utilize them in grading, although the amount of weight to be given to the results was not specified. Ss were instructed in the system of scoring to be used as follows:

---

[1] Permission to use this test was granted by the Educational Testing Service.

003 548

Each of the questions in this test is followed by suggested answers. Assign a number from 0 to 9 to each suggested answer, depending on how strongly you feel that the answer is correct. If you believe that only one suggested answer is correct, mark that answer with a 9 and mark the other(s) with zeros. If you like the suggested answers equally, assign the same number to each. The sum of the three responses should add up to 9 . . .

If your answer is closer to the right answer, you will get a positive score. If it is closer to the wrong answer you will get a negative score. The scores vary from -1 to +1. They are multiplied by your certainty, (C).

The test itself was preceeded by a six-item practice test at the end of which subjects were given the right answer for each question and could ask any question about the instructions. They were told that for the practice test there was one single right answer, but for the test itself, there might or might not be more than one single right answer to each item. The items were scored using the Weighted Euclidean function $S = C(1 - 2D / D_{max})$ where:

C = Confidence ($0 \leq C \leq 9$)
D = Distance from S's response to the criterion group response.
D_{max} = Maximum distance attainable from the criterion group response.

Ss were asked to fill out a personal data sheet, and were given a test of 5 personality variables.[2/] From these instruments the following variables were measured:

1. Sex: Male = 1, Female = 2
2. Year in School: 1 = Sophomore, 2 = Junior
3. Score: Mean weighted Euclidean score on the 15 item writing test
4. Attitude: 0 = maximum dislike for test, 9 = maximum liking
5. Confidence: 0 = minimum confidence in responses, 9 = maximum
6. Autonomy: Scale score from Personality Research Inventory
7. Harm Avoidance: Personality Research Inventory
8. Impulsivity: Scale from Personality Research Inventory
9. Order: Scale from Personality Research Inventory
10. Succorance: Scale from Personality Inventory
11. Social Class: (on a three-point scale) Low = 1, Middle = 2, Upper = 3
12. Appropriateness of Confidence (WPLN)
13. Propensity to gamble (PLN)
14. Appropriateness of Confidence on an item of medium difficulty
15. Gambling propensity on an item of medium difficulty
16. Appropriateness of Confidence on an easy item, #7
17. Gambling propensity on an easy item
18. Appropriateness on a difficult item, #13
19. Gambling propensity on a difficult item

2/ Scales Au, Ha, Im, Or, Su, from Douglas Jackson, Personality Research Form, Form AA, Research Psychologists Press, Inc. 1965.

Some explanation is necessary on the computation of variables 12 through 19.

The propensity to gample, PLN, for an item was equal to the sum of the squares of the differences between numerical response for each of the responses and three, divided by six. That is:

$$PLN = \left(\sum_{j=1}^{3} (r_j - 3)^2\right)/6 \text{ for the } i^{th} \text{ item,}$$

where $0 \leq r_j \leq 9$

and $\sum_{j=1}^{3} r_j = 9$, j = option number

Since subject responses ranged from 0 to 9 for the three options, Ss who had no preference for the options, and who expressed this lack of preference by responding (3,3,3) to the three options would receive a PLN equal to zero. On the other hand, S showing a complete preference for a single option (propensity to gamble) would receive PLN = (36 + 9 + 9)/6 = 9. Thus PLN is an index of the subject's tendency to select a single option. PLN for a test would then consist of the average value of PLN over all the items.

Appropriateness of confidence compares S's PLN with his expressed confidence in the item. For the i^{th} item, appropriateness of confidence (WPLN) is the absolute value of the difference between S's PLN for that item and his confidence measure, C_i:

$$WPLN_i = \left|PLN_i - C_i\right|$$

Theoretically, a person with no knowledge should declare $C_i = 0$ and distribute his responses (3,3,3). This would make PLN = 0 and C = 0. Thus a score of 0 on WPLN indicates congruence between PLN and C_i. A S who is certain of his response would mark one option with a nine and the other options with zeroes. This would make PLN = 9. If he was that certain, he should also mark C = 9, again giving WPLN = 0. Positive values of WPLN indicate a discrepancy between confidence and one's behavior in distributing his responses.

Means and standard deviations of the 19 variables under the relaxed and the incentive conditions are shown in Tables 1 and 2.

The reliability of the test under the incentive condition was 0.261. Under the relaxed condition it was 0.493. Although the mean scores were significantly higher under the incentive condition, the reliability of these scores was consistently lower. Although these reliabilities may seem low, it must be remembered that the items were only 1/4 of the items from the original test. When corrected for length, reliabilities are close to the published values.

Ss reported a slightly more favorable attitude toward the test under the incentive condition. Although the average liking in both was low there was a significantly greater amount of confidence than there was in the relaxed group, along with a significantly higher propensity to improve their score. The

confidence expressed in the incentive group was more congruent with their distribution of preference than was the confidence expressed by the relaxed group. Confidence was most appropriate on the easy item, and was least appropriate on the item of moderate difficulty.

Using data shown in Table 3, Grozelier (1970) concluded that girls were slightly more sensitive to the incentive effect than were boys. With regard to the level of risk-taking, boys were rather conservative and girls high-risk oriented. This would follow from an assumption that the motive to achieve success would be stronger among boys whereas girls would rather be failure avoidance oriented.

TABLE 1 - INCENTIVE GROUP

	VARIABLE	MEAN	ERROR	N	ST.DEV.	ERROR
1	Sex	1.545	0.048	110	0.500	0.
2	Year	1.664	0.045	110	0.475	0.015
3	Score	4.344	0.306	110	3.209	0.197
4	Attitude	3.336	0.230	110	2.409	0.123
5	Confidence	7.912	0.085	110	0.891	0.089
6	Autonomy	9.229	0.345	109	3.597	0.193
7	Harm Avoidance	7.321	0.299	109	3.118	0.179
8	Impulsivity	10.376	0.291	109	3.036	0.181
9	Order	10.000	0.370	109	3.866	0.237
10	Succorance	9.899	0.381	109	3.979	0.198
11	Social Class	1.764	0.062	110	0.649	0.047
12	Approp. Confidence	1.527	0.071	110	0.739	0.053
13	Propensity to Gamble	7.130	0.092	110	0.963	0.069
14	Approp. on Med. Diff.	2.447	0.222	110	2.328	0.119
15	Gamble on Med. Diff.	6.571	0,319	110	3.349	0.117
16	Approp. on Easy Item	0.292	0.090	110	0.942	0.232
17	Gamble on Easy Item	8.890	0.070	110	0.730	0.283
18	Approp. on Hard Item	2.174	0.218	110	2.291	0.164
19	Gamble on Hard Item	6.346	0.298	110	3.126	0.150

TABLE 2 - RELAXED GROUP

	VARIABLE	MEAN	ERROR	N	ST.DEV.	ERROR
1	Sex	1.477	0.048	111	0.502	0.
2	Year	1.631	0.046	111	0.485	0.012
3	Score	3.680	0.344	111	3.619	0.229
4	Attitude	3.027	0.244	111	2.574	0.122
5	Confidence	7.324	0.149	111	1.570	0.232
6	Autonomy	9.410	0.355	105	3.642	0.185
7	Harm Avoidance	7.571	0.355	105	3.642	0.198
8	Impulsivity	11.448	0.298	105	3.051	0.186
9	Order	8.667	0.415	105	4.251	0.263
10	Succorance	9.371	0.349	105	3.574	0.253
11	Social Class	1.982	0.067	110	0.704	0.063
12	Approp. Confidence	1.638	0.119	111	1.256	0.175
13	Propensity to Gamble	6.586	0.116	111	1.218	0.095
14	Approp. on Med. Diff.	2.432	0.239	111	2.521	0.176
15	Gamble on Med. Diff.	5.414	0.328	111	3.454	0.120
16	Approp. on Easy Item	0.383	0.142	111	1.501	0.354
17	Gamble on Easy Item	8.482	0.186	111	1.960	0.358
18	Approp. on Hard Item	2.220	0.218	111	2.296	0.216
19	Gamble on Hard Item	5.241	0.329	111	3.469	0.115

TABLE 3 - GROUP MEANS

		TOTAL	SEX		GRADE		SOCIAL CLASSES		
			M	F	SO	JU	I	II	III
PLN MEAN	NEUTRAL CONDITION	6.5	6.4	6.7	6.4	6.6	6.2	6.6	6.9
	INCENTIVE CONDITION	7.0	6.9	7.2	7.1	7.0	7.1	7.0	7.8
	TOTAL	6.8	6.6	6.9	6.7	6.8	6.8	6.8	7.3
PLN ITEM 7 (Easy)	NEUTRAL CONDITION	8.5	8.6	8.4	8.7	8.4	8.7	8.6	8.2
	INCENTIVE CONDITION	8.9	8.9	8.9	8.8	8.9	8.9	9.0	8.3
	TOTAL	8.7	8.7	8.7	8.7	8.7	8.8	8.7	8.3
PLN ITEM 6 (Average)	NEUTRAL CONDITION	5.5	5.6	5.3	6.3	5.1	5.2	5.5	5.1
	INCENTIVE CONDITION	5.8	5.0	6.3	5.1	6.1	5.1	5.9	6.1
	TOTAL	5.6	5.4	5.9	5.7	5.6	5.1	5.7	5.6
PLN ITEM 13 (Difficult)	NEUTRAL CONDITION	5.4	4.5	6.3	5.8	5.0	4.6	5.5	5.4
	INCENTIVE CONDITION	6.1	5.4	6.6	6.4	6.0	6.9	6.0	5.9
	TOTAL	5.7	4.9	6.4	6.1	5.5	6.1	5.7	5.7

Social Class

On item 6, higher class subjects appear to be the most conservative. This was particularly conspicious under the incentive condition (PLN_i mean = 5.1 for the higher class, PLN_i mean = 5.9 for the middle class, PLN_i mean = 6.1 for the lower class).

Middle class subjects appeared as moderate risk takers and appeared as motivated to achieve success whereas lower class were "fear of failure" oriented.

Middle class students received slightly higher scores than the two other classes (though not statistically significant). They tended to display a motivation to achieve success.

Lower class students fared the worst on this test. They were most risk minded and therefore obtained the lowest scores because confidence testing penalizes guessing and rewards the acknowledgement of partial knowledge.

Correlations were computed for each of the two samples for all 19 variables. The correlation matrices are shown in Tables 4 and 5. Correlations larger than $r = .195$ will be examined. For a single pair of variables, a correlation of 0.195 indicates a significant departure from 0.0 at the 0.025 level with 100 degrees of freedom. (Walker and Lev, 1953). Comparing significant correlations in the two matrices, it can be seen that there was a significant relationship between sex and attitude toward the test with the girls liking it better than the boys. This sex difference was accentuated under the incentive condition. The males were more Autonomous and less Succorant in both groups. This should be expected because the personality test was not involved in the incentive instructions. Finally, only the difficult item provided a significant correlation with appropriateness of judgment of confidence and the propensity to gamble with the females showing a greater willingness to make extreme choices. and also exhibiting greater congruence between their feelings of certainty and their behavior in responding to the items. That is, the females were more inclined to chose single responses, but they also felt more certain about their choices than did the males. Confidence was significantly related to score under both conditions, though the relationship was higher under the relaxed condition. That is, subjects were more willing to take extreme positions under the relaxed condition.

TABLE 4

CORRELATION COEFFICIENTS - RELAXED CONDITION

	VARIABLE	1	2	3	4	5	6	7	8	9	10
1	Sex	1.000									
2	Year	0.022	1.000								
3	Score	0.096	-0.039	1.000							
4	Attitude	0.278	-0.036	0.110	1.000						
5	Confidence	0.194	0.062	0.209	0.230	1.000					
6	Autonomy	-0.283	-0.211	-0.039	-0.183	-0.189	1.000				
7	Harm Avoidance	-0.090	0.074	-0.102	-0.004	-0.028	-0.423	1.000			
8	Impulsivity	0.066	0.101	0.044	-0.041	-0.002	0.168	-0.357	1.000		
9	Order	0.085	-0.018	-0.133	-0.016	0.099	-0.306	0.314	-0.543	1.000	
10	Succorance	0.396	0.066	0.103	0.365	0.218	0.539	0.241	-0.058	0.089	1.000
11	Social Class	0.076	0.194	-0.058	0.156	0.023	-0.105	-0.016	-0.065	-0.039	0.070
12	Approp. Confidence	-0.033	-0.190	-0.170	-0.188	-0.325	0.089	0.029	-0.089	0.034	-0.065
13	Propensity to Gamble	0.134	0.085	0.221	0.137	0.180	-0.071	0.008	-0.016	0.045	0.047
14	Approp. on Med. Diff.	-0.006	0.059	-0.028	-0.078	-0.205	0.053	0.018	0.133	-0.080	0.059
15	Gamble on Med. Diff.	0.076	-0.235	0.082	0.125	0.122	0.192	-0.023	-0.196	0.057	-0.070
16	Approp. on Easy Item	-0.119	-0.154	-0.114	-0.206	-0.547	0.095	0.122	-0.105	0.045	-0.093
17	Gamble on Easy Item	0.001	-0.117	0.219	0.137	0.293	-0.133	0.021	-0.080	0.043	0.228
18	Approp. on Hard Item	-0.202	-0.119	-0.036	-0.089	-0.254	0.069	0.037	0.077	0.012	-0.045
19	Gamble on Hard Item	0.270	-0.122	0.178	0.199	0.123	-0.030	0.031	-0.112	0.067	0.076

		11	12	13	14	15	16	17	18	19
11	Social Class	1.000								
12	Approp. Confidence	0.036	1.000							
13	Propensity to Gamble	0.176	-0.481	1.000						
14	Approp. on Med. Diff.	-0.060	0.370	-0.251	1.000					
15	Gamble on Med. Diff.	0.046	0.026	0.218	-0.695	1.000				
16	Approp. on Easy Item	-0.070	0.597	-0.178	0.278	0.093	1.000			
17	Gamble on Easy Item	-0.159	-0.036	-0.009	0.080	-0.051	-0.153	1.000		
18	Approp. on Hard Item	-0.063	0.343	-0.217	0.263	-0.012	0.342	0.167	1.000	
19	Gamble on Hard Item	0.036	0.034	0.297	-0.013	0.194	0.080	-0.065	-0.458	1.000

TABLE 5.

CORRELATION COEFFICIENTS - INCENTIVE CONDITION

	VARIABLE	1	2	3	4	5	6	7	8	9	10
1	Sex	1.000									
2	Year	-0.148	1.000								
3	Score	0.150	-0.219	1.000							
4	Attitude	0.334	0.188	0.167	1.000						
5	Confidence	0.100	-0.159	0.178	0.192	1.000					
6	Autonomy	-0.284	0.029	-0.060	-0.173	-0.045	1.000				
7	Harm Avoidance	0.116	-0.007	0.032	0.058	-0.080	-0.206	1.000			
8	Impulsivity	0.142	-0.007	-0.041	0.065	-0.007	0.056	-0.382	1.000		
9	Order	0.024	0.115	0.093	-0.019	-0.046	-0.080	0.260	-0.374	1.000	
10	Succorance	0.429	-0.018	0.143	0.167	-0.053	-0.678	0.358	0.027	0.072	1.000
11	Social Class	0.033	0.097	0.093	0.139	0.073	0.090	0.115	-0.057	0.124	0.051
12	Approp. Confidence	-0.095	0.013	-0.074	-0.030	-0.398	-0.009	0.150	-0.077	0.178	0.081
13	Propensity to Gamble	0.158	0.024	0.203	0.215	0.473	-0.066	-0.191	0.152	-0.158	-0.013
14	Approp. on Med. Diff.	-0.199	-0.110	-0.034	-0.152	0.013	-0.039	-0.012	-0.139	0.036	-0.034
15	Gamble on Med. Diff.	0.160	0.159	0.087	0.174	0.174	0.120	0.024	0.106	-0.049	-0.045
16	Approp. on Easy Item	-0.074	-0.074	0.097	-0.050	0.108	0.178	0.137	-0.102	-0.081	-0.209
17	Gamble on Easy Item	-0.003	0.141	-0.050	-0.007	-0.138	-0.061	-0.114	0.080	0.079	0.123
18	Approp. on Hard Item	-0.244	0.105	0.166	-0.080	-0.300	0.077	0.155	-0.056	0.219	0.040
19	Gamble on Hard Item	0.253	-0.104	-0.273	0.063	0.033	0.001	-0.211	0.096	-0.108	-0.039

11	Social Class	1.000									
12	Approp. Confidence	-0.059	1.000								
13	Propensity to Gamble	0.136	-0.707	1.000							
14	Approp. on Med. Diff.	-0.023	0.138	-0.157	1.000						
15	Gamble on Med. Diff.	0.139	-0.165	0.304	-0.807	1.000					
16	Approp. on Easy Item	0.174	-0.053	0.047	0.247	-0.049	1.000				
17	Gamble on Easy Item	-0.133	0.209	-0.136	-0.101	0.091	-0.710	1.000			
18	Approp. on Hard Item	0.131	0.443	-0.226	0.063	-0.076	0.067	0.027	1.000		
19	Gamble on Hard Item	-0.099	-0.189	0.214	-0.037	0.008	-0.121	0.060	-0.567	1.000	

Thus the subjects seemed to be more motivated by fear of failure than by potential reward. It is of additional interest to note that there was no relationship between score and the gamble score on the easy items in the incentive condition, while the significant relationship was on the hard item under the incentive condition. In fact, the gamble hard score - score correlation changed sign going from the relaxed condition to the incentive condition. That is, for the high scoring Ss, there was a tendancy to assume extreme positions on the hard items under the relaxed condition, but an unwillingness to do so under the incentive condition. That is, where grades were at stake, the high scoring Ss played the cautious role. S's attitude toward the test was related primarily to his confidence, although there was also a significant relationship with the gamble score in the incentive condition. Confidence was significantly related to inappropriatness of judgment and to willingness to take extreme positions under the incentive condition. That is, under the incentive, subjects who were confident about their responses were more willing to take extreme positions in responding. However, these extreme positions did not match their degrees of confidence very well. Several other of the item scores were related to confidence in the relaxed condition, while the gamble score became less important. The personality variables showed substantial intercorrelations as did the cluster of gamble and appropriatness scores. The significant negative correlations between the gamble and the appropriatness scores is due to the fact that these two scores are not independent of one another. The negative sign becomes obvious when one examines the means of computation of the appropriatness score (WPLN) from gamble score (PLN).

In order to better understand what variables contributed to S's expression of confidence, a regression analysis was performed. No significant regression held between confidence and any other variables, although high succorance and low harm avoidance did contribute a small amount to the prediction of confidence in the relaxed condition only.

Seventeen of the scores were factor analyzed. The PLN and WPLN variables for the item of medium difficulty were left out since they did not seem to provide much information. A principal components analysis was first performed. Then the principal components were rotated according to the following specifications: A maximum of nine factors were to be extracted, the lower limit of eigenroots was set at 1.00 and no factors were to have loadings of less than .30 for at least one variable. According to these specifications, seven factors were rotated. Ten rotations were required in the incentive condition. Thirteen were required in the relaxed condition. The factor matrix is shown in Tables 6 and 7 Loadings in excess of 0.30 are underlined.

In interpreting these results, it should be recalled that a low numerical score on the Appropriate variable means that a person's responses were congruent with his confidence. The factor analysis did not reveal much about confidence, except to underline the fact that there is a dependence between it and the gamble and appropriate measures. This is illustrated in Factor 1 in both conditions. Factor 2 is made up of sex and several personality variables. Attitude is also a relevant variable

TABLE 6

FACTOR MATRIX--INCENTIVE CONDITION

			1	2	3	4	5	6	7
Factor Number			1	2	3	6	5	4	7
Factor Number (before Rotation)									
Sum of Squares			2.245	2.069	1.847	1.845	1.687	1.621	1.302
PerCent 17 Factors			12.5	24.0	34.2	44.5	53.8	62.9	70.1
7 Factors			17.8	34.2	48.8	63.5	76.8	89.7	100.0
No.	Name	Communality							
1	Sex	0.741	-0.033	0.463	-0.006	0.393	-0.070	0.485	0.364
2	Year	0.802	0.000	-0.002	-0.125	-0.159	0.034	0.278	-0.826
3	Score	0.733	0.250	0.087	-0.025	-0.461	0.049	0.230	0.629
4	Attitude	0.511	0.081	0.260	-0.004	0.069	-0.071	0.652	-0.023
5	Confidence	0.572	0.699	-0.025	0.054	0.057	0.071	0.113	0.244
6	Autonomy	0.806	-0.107	-0.877	0.093	-0.005	-0.061	0.113	0.025
7	Harm Avoidance	0.604	-0.195	0.390	0.257	-0.088	-0.569	0.124	-0.022
8	Impulsivity	0.738	-0.028	0.022	-0.100	0.004	-0.837	0.157	0.022
9	Order	0.619	-0.101	0.005	-0.176	-0.114	0.743	0.108	0.023
10	Succorance	0.825	-0.073	0.888	-0.104	-0.051	0.039	0.112	0.062
11	Social Class	0.419	0.044	-0.080	0.254	-0.165	0.179	0.527	-0.096
12	Approp. Confidence	0.772	-0.840	0.008	-0.104	-0.198	0.097	0.004	0.082
13	Propensity to Gamble	0.793	0.833	0.042	0.057	0.057	-0.173	0.245	-0.022
14	Approp. on Easy Item	0.827	0.022	-0.156	0.888	-0.091	0.023	0.031	0.060
15	Gamble on Easy Item	0.830	-0.164	0.012	-0.894	-0.002	0.003	0.028	-0.049
16	Approp. on Hard Item	0.746	-0.365	-0.030	0.019	-0.778	0.065	0.043	-0.028
17	Gamble on Hard Item	0.762	0.063	-0.046	-0.086	0.860	-0.086	0.039	0.009

TABLE 7

FACTOR MATRIX--RELAXED CONDITION

			1	2	3	4	5	6	7
Factor Number			1	5	2	6	3	7	4
Factor Number (before rotation)									
Sum of Squares			2.233	2.128	2.055	1.623	1.545	1.415	1.313
PerCent 17 Factors			12.4	24.2	35.6	44.7	53.2	61.1	68.4
7 Factors			18.1	35.4	52.1	65.3	77.8	89.3	100.0
No.	Name	Communality							
1	Sex	0.622	0.051	0.671	-0.115	0.381	0.072	-0.050	-0.056
2	Year	0.610	0.124	0.019	-0.010	-0.014	-0.705	0.073	-0.302
3	Score	0.651	0.068	0.172	-0.179	0.034	0.101	0.706	0.273
4	Attitude	0.562	0.257	0.578	-0.051	0.009	0.287	0.072	-0.269
5	Confidence	0.633	0.719	0.285	0.059	-0.048	0.097	0.063	0.124
6	Autonomy	0.726	-0.040	0.608	-0.439	-0.032	0.389	-0.094	0.022
7	Harm Avoidance	0.638	0.179	0.133	0.686	0.063	-0.306	0.129	0.049
8	Impulsivity	0.719	0.001	0.038	-0.802	0.037	-0.262	0.032	0.030
9	Order	0.664	0.072	0.047	0.790	-0.059	0.078	-0.145	0.034
10	Succorance	0.722	0.045	0.815	0.141	0.059	0.161	0.073	0.033
11	Social Class	0.697	0.011	0.159	-0.011	0.037	0.024	0.005	-0.818
12	Approp. Confidence	0.736	0.710	0.101	0.015	0.101	0.227	-0.201	0.036
13	Propensity to Gamble	0.727	0.237	-0.034	0.081	-0.236	0.016	0.[illegible]6	-0.307
14	Approp. on Easy Item	0.812	0.886	0.066	0.097	0.028	0.107	0.0[illegible]7	0.044
15	Gamble on Easy Item	0.645	0.276	0.300	0.082	0.461	0.204	0.171	0.434
16	Approp. on Hard Item	0.775	-0.387	-0.013	-0.034	0.783	0.104	-0.011	0.008
17	Gamble on Hard Item	0.764	-0.111	0.212	0.070	-0.748	0.253	0.271	0.071

in the relaxed condition, but the importance of attitude in this factor is much reduced in the incentive condition. Factor 3 in the relaxed condition and Factor 5 in the incentive condition are quite similar and are made up entirely of the personality factors. Factor 4 is perhaps the only one of much interest. It shows a relationship among sex and the way in which Ss deal with the difficult items. This, however, only confirms what has been previously said about sex differences with respect to making dogmatic choices on items.

Conclusions

The findings of this study are summarized as follows:

1. Under incentive conditions, scores on confidence tests are higher, and reliability significantly lower when compared to the relaxed condition.

2. Females have a greater tendency toward taking extreme positions than males, especially in the incentive condition.

3. Subjects in the incentive group liked the test better, had more of a tendency to take extreme positions, and made more appropriate estimates of their confidence.

4. Middle SES subjects, compared to both upper and lower SES subjects, made higher scores and more appropriate estimates of confidence. They seemed to be motivated more by desire for success than fear of failure.

5. High scoring subjects gambled more on difficult items under the relaxed condition, but gambled less on difficult items in the incentive condition.

6. Liking of tests was directly related to confidence.

7. There was no significant regression between confidence and the battery of personality variables, although high succorance and low harm avoidance made small contributions to prediction.

Much work remains to be done in studying confidence testing. Although it is clear that technical improvements may be made in the reliability and validity of tests through confidence scores, it is also clear that subjects do not handle their confidence uniformly. What is confidence to one may be hazard to another. As Wang and Stanley state, (1970)

> "The derivation of optimum response strategies in multiple choice testing represents an application of mathematical decision theory which underscores the decision process inherent in such tests. The success of testing procedures which attempt to control the decision process will be critically dependent on the ability of subjects to effectively use optimal strategies. It is not certain that all subjects are equally capable of learning to use such strategies."

Understanding optimal strategies of probability assessment is likely to be the most significant outcome of further research on confidence testing. Although Bruner (1956) pointed out two basic differences in the way subjects use their

confidences - the sentry condition and accuracy condition, and demonstrated empirical evidence of these two modes of behavior, there are other complex conditions which intervene between a subjective probability and a decision or action. Since it is possible, although not guaranteed that one may assess subjective probabilities accurately by means of reproducing scoring functions, some basic research steps are needed. First, subjects in experiments need experience in utilizing confidence testing. It takes awhile to learn to respond intelligently to the rules of that game. Second, the possibility of applying the relative operating characteristic to confidence testing needs to be explored (Swets, 1973). Once a more valid interpretation of subjective probabilities was available, further study might be made of the use of optimal strategies by subjects in problematic situations. Such strategies would perhaps start with what is known about optimal search procedures in polychotomic trees (Watanabe, 1969).

A sizeable field in this area remains unplowed. How do students react to problematic situations? Are students able to assess their state of information and respond intelligently to it? Do our teaching and testing practices make them aware that there are differences among the ways we use our information? And to repeat DeFinnetti, "How can we become better probability assessors?"

REFERENCES

Bruner, J. S., Goodnow, J. J., and Austin, G. A., A Study of Thinking. New York: Wiley & Sons, 1956.

DeFinetti, B., Methods for Discriminating Levels of Partial Knowledge Concerning a Test Item, British Journal of Mathematical and Statistical Psychology, 1965, 18, 87-123.

Grozelier, A. M., The Influence of Social Values on Risk-Taking in a Confidence Testing Situation. M. A. Dissertation, University of Chicago, Chicago, Illinois, 1970.

Hambelton, R. K., Roberts, D. M., and Traub, R. E., A Comparison of the Reliability and Validity of Two Methods for Assessing Partial Knowledge on Multiple Choice Tests. Journal of Educational Measurement, 1970, 7, 75-90.

Rippey, R. M., A Comparison of Five Different Scoring Functions for Confidence Tests. Journal of Educational Measurement, 7, No. 3, Fall 1970, 165-170.

Shuford, E., Albert, A., and Massengill, N. E., Admissable Probability Measurement Procedures. Psychometrika, 1966, 31, 125-145.

Swets, J. A., The Relative Operating Characteristics in Psychology, Science, 7 December 1973, pp. 990-1000.

Walker, H. M., and Lev, J., Statistical Inference. New York: Holt, Rinehard, and Winston, 1953.

Wang, M. W., and Stanley, J. C., Differential Weighting: A Review of Methods and Empirical Studies, Review of Educational Research, 1970, 40, 663-705.

Watanabe, S., Knowing and Guessing: A Formal and Quantitative Study. New York: Wiley & Sons, 1969.